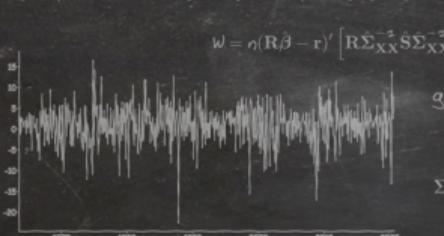


$$KS = \max_{\tau} \left| \sum_{t=1}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{z}{\sqrt{n}} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

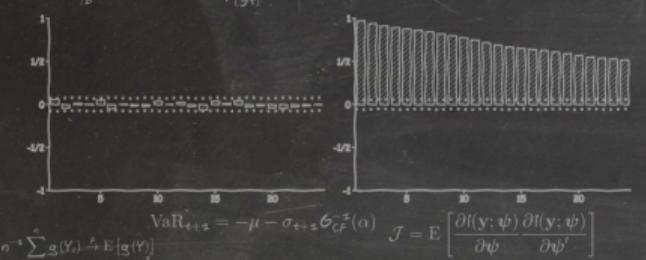
Introduction

VaR, ES and Density Forecasting

$$\begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix} = \begin{bmatrix} \pi_{x_2} \\ \pi_{y_2} \end{bmatrix} + \begin{bmatrix} \alpha_x e_i \\ \alpha_y e_i \end{bmatrix} + \pi_{\pi} \begin{bmatrix} \Delta x_{t-2} \\ \Delta y_{t-2} \end{bmatrix} + \dots + \pi_{\pi} \begin{bmatrix} \Delta x_{t-k} \\ \Delta y_{t-k} \end{bmatrix} + \begin{bmatrix} \eta_{x,i} \\ \eta_{y,i} \end{bmatrix}$$



$$\rho_s = \frac{\gamma_s}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])]}{V[y_t]} \Rightarrow -2X'y + 2X'X\beta = \sigma \quad -2X'y - 2X'\hat{\epsilon} = \sigma \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y,x}$$



$$VaR_{t+2} = -\mu - \sigma_{t+2} G_{C^2}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial J(y; \psi)}{\partial \psi} \frac{\partial J(y; \psi)}{\partial \psi'} \right]$$

$$\Delta y_t = \phi_\theta + \delta_z t + \gamma y_{t-z} + \sum_{\rho=1}^p \phi_\rho \Delta y_{t-\rho} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \quad BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$t = \frac{\sqrt{n} (R\hat{\theta} - r)}{\sqrt{RG^{-1}\Sigma(G^{-1})'R'}} \xrightarrow{d} N(\theta, 1) \quad N(\mu_1 + \beta'(x_2 - \mu_2), \Sigma_{22} - \beta'\Sigma_{21}\beta)$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\mathcal{Z}^4] \quad \arg\min_{\beta} (y - X\beta)(y - X\beta) + \sum_{j=1}^k |\beta_j|$$

$$\hat{\ell}^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \hat{\ell}^0(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{1-\rho}(1 - \rho)^{\rho-1}}{B(\alpha, \beta)} \quad E \left[\left(\beta (1 + r_{t+s}) \left(\frac{U(c_{t+s})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho$$

$$Y_t = \delta + Y_{t-z} + \epsilon_t \quad \lambda(y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i!) \quad \hat{\Gamma}_i = \frac{l+1-i}{l+1} (\bar{\Gamma}_i + \bar{\Gamma}_i')$$

$$\Sigma = \begin{bmatrix} \Sigma_{zz} & \Sigma_{z2} \\ \Sigma'_{z2} & \Sigma_{22} \end{bmatrix} \quad Y_i = \beta_2 X_i + \beta_2 X_i I_{[X_i > \kappa]} + \epsilon_i \quad \sqrt{\tau}(R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right)$$

$$\text{Power}(\mu_1, \mu_2, \sigma, \alpha, N) = \Phi \left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma}} \right) + \alpha - \Phi \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma}} \right) \quad \frac{\partial l(\theta, y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2} \quad A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$$\frac{\partial l(\theta, y)}{\partial \sigma^2} = -\frac{\alpha}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^4} \quad \hat{\ell}^0(x_1 | x_2 \in B) = \frac{\int_B \hat{\ell}^0(x_1, x_2) dx_2}{\int_B \hat{\ell}^0(x_2) dx_2} \quad \lambda_{trace}(r) = -\tau \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$z_t = \Upsilon z_{t-2} + \xi_t \quad \sigma_t^2 = \omega + \alpha V_{t-2}^2 + \beta \sigma_{t-2}^2 \quad \Sigma_t = CC' + AA' \odot \epsilon_{t-2} \epsilon_{t-2}' + BB' \odot \Sigma_{t-2}$$

Risk Measurement Overview

- What is risk?
- What is Value-at-Risk?
- How can VaR be measured and modeled?
- How can VaR models be tested?
- What is Expected Shortfall?
- How can densities be forecasted?
- How can density models be evaluated?
- What is a coherent risk measure?

Risk

- What is risk?
 - ▶ Market Risk
 - ▶ Liquidity Risk
 - ▶ Credit Risk
 - ▶ Counterparty Risk
 - ▶ Model Risk
 - ▶ Estimation Risk
- Today's focus: **Market Risk**
- Tools
 - ▶ Value-at-Risk
 - ▶ Expected Shortfall
 - ▶ Density Estimation

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Defining Value-at-Risk

VaR, ES and Density Forecasting

$$\Delta y_t = \phi_\theta + \delta_{\pi} t + \gamma y_{t-\pi} + \sum_{\rho=1}^p \phi_\rho \Delta y_{t-\rho} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \\ BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$t = \frac{\sqrt{n}(\hat{R}\theta - r)}{\sqrt{RG^{-1}\Sigma(G^{-1})'R'}} \xrightarrow{d} N(\mu, 1)$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\mathcal{Z}^4] \quad N(\mu_1 + \beta'(x_2 - \mu_2), \Sigma_{22} - \beta'\Sigma_{21}\beta)$$

$$\hat{\ell}^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \text{argmin}_{\beta} (y - X\beta)(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j|$$

$$\frac{\Delta x_i}{\Delta y_0} = \frac{\pi_{\rho_2}}{\pi_{\rho_1}} + \frac{\alpha_1}{\alpha_2} \epsilon_1 + \pi_1 \left[\frac{\Delta x_{i-1}}{\Delta y_{i-1}} \right] + \dots + \pi_0 \left[\frac{\Delta x_{i-p}}{\Delta y_{i-p}} \right] + \left[\eta_{i+1} \right]$$

$$\hat{\ell}^1(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{1-\pi}(1 - \rho)^{\beta-\pi}}{B(\alpha, \beta)} \quad E \left[\left(\beta(1 + r_{t+\pi}) \left(\frac{U(c_{t+1})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho$$

$$\omega = \eta(R\beta - r)' \left[R \hat{\Sigma}_{XX} \hat{S} \hat{\Sigma}_{XX}^{-1} R' \right]^{-1} (R\beta - r) \xrightarrow{d} \chi_m^2 \quad Y_t = \delta + Y_{t-\pi} + \epsilon_t$$

$$g(e) = \frac{1}{\tau_n} \sum_{t=1}^{\tau} K \left(\frac{\hat{e}_t - e}{h} \right) \quad l(\lambda, y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i)$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \hat{S}^{\text{EW}} = \hat{\Gamma}_\theta + \sum_{i=1}^l \frac{l+\pi - i}{l+\pi} \left(\hat{\Gamma}_i + \hat{\Gamma}'_i \right) \quad Y_i = \beta_1 X_i + \beta_2 X_i I_{[X_i > \kappa]} + \epsilon_i$$

$$\sqrt{\tau}(R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right) \quad \text{Power}(\mu_1, \mu_2, \sigma_1, \alpha, N) = \Phi \left(\frac{\mu_2 - \mu_1}{\sqrt{N}} \right) + \alpha - \Phi \left(\frac{\mu_1 - \mu_2}{\sqrt{N}} \right)$$

$$\rho_\pi = \frac{\gamma_\pi}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-\pi} - E[y_{t-\pi}])]}{V[y_t]} \Rightarrow -2X'y - 2X'X\beta = \nu \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y, x}$$

$$\beta \cap \beta = \{x : x \in \beta \text{ and } x \in \beta\} \quad \frac{\partial l(\theta, y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2}$$

$$A \cap B = \{x : x \in A \text{ and } x \in B\} \quad \lambda_{\text{trace}}(r) = -\tau \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f(x_2) dx_2}$$

$$z_t = \Upsilon z_{t-\pi} + \xi_t \quad \sigma_t^2 = \omega + \alpha V_{t-\pi}^2 + \beta \sigma_{t-\pi}^2$$

$$\Sigma_t = CC' + AA' \odot \epsilon_{t-\pi} \epsilon_{t-\pi} + BB' \odot \Sigma_{t-\pi}$$

$$\text{VaR}_{t+\pi} = -\mu - \sigma_{t+\pi} G_{C'}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial l(y; \psi)}{\partial \psi} \frac{\partial l(y; \psi)}{\partial \psi'} \right]$$

$$\sum_{i=1}^n g_i(Y_i) \xrightarrow{d} E[g(Y)]$$

Value-at-Risk

- Value-at-Risk is a standard tool of risk management
 - ▶ Basel Accord

Definition (Value-at-Risk)

The α Value-at-Risk (VaR) of a portfolio is defined as the largest number such that the probability that the loss in portfolio value over some period of time is greater than the VaR is α ,

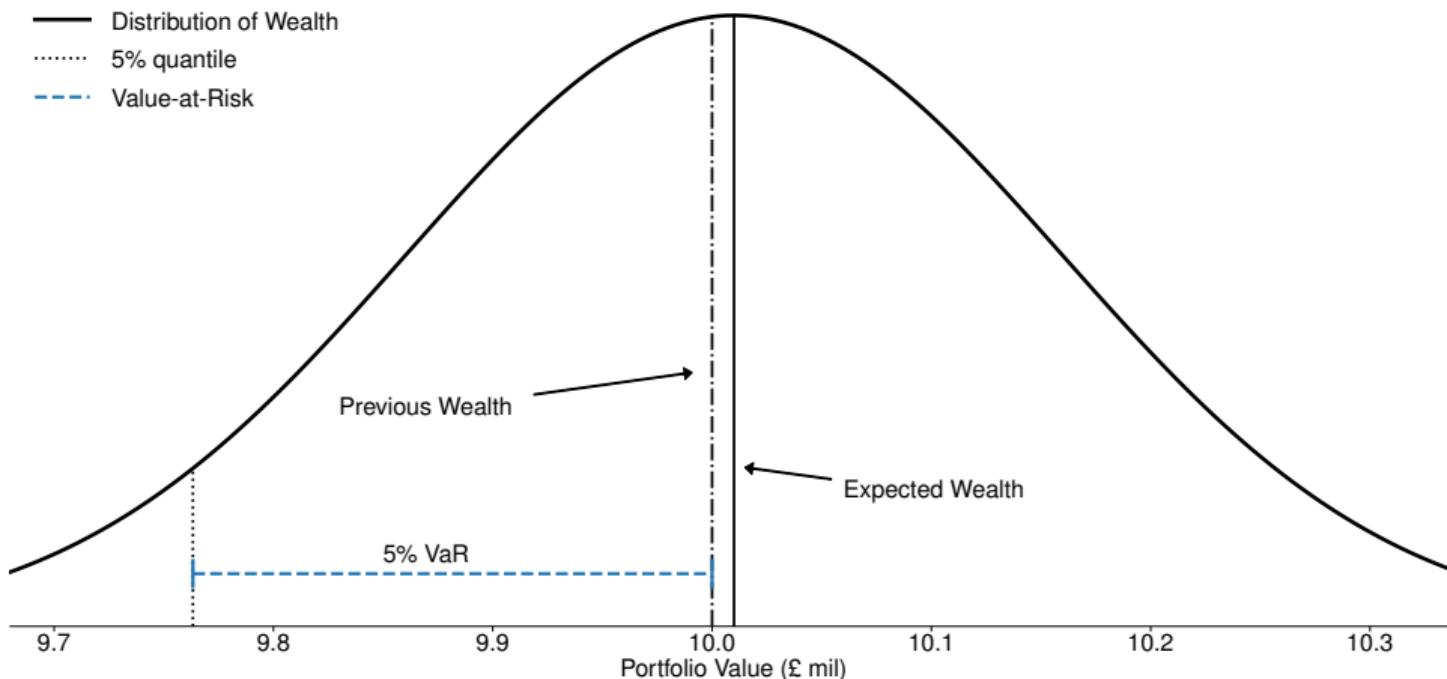
$$Pr(R < -VaR) = \alpha$$

where $R = W_1 - W_0$ is the total return on the portfolio, W_t , $t = 0, 1$, is the value of the assets in the portfolio and 1 and 0 measure an arbitrary time span (e.g. one day or two weeks).

- Units are \$, £, ¥
- Almost always *positive*
- It is a **quantile**

Value-at-Risk in a picture

- Returns are $N(.001, .015^2)$
- W_0 is £10,000,000



$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Alternative Value-at-Risk Measures

VaR, ES and Density Forecasting

$$\Delta y_t = \phi_\theta + \delta_z t + \gamma y_{t-z} + \sum_{p=1}^p \phi_p \Delta y_{t-p} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \\ BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\epsilon^4] \quad N(\mu_1 + \beta' (x_2 - \mu_2), \Sigma_{22} - \beta' \Sigma_{21} \beta) \\ \hat{\ell}^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \arg\min_{\beta} (y - X\beta)(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j| \\ \hat{\ell}^1(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{1-\rho}(1 - \rho)^{\beta-1}}{B(\alpha, \beta)} \quad E \left[\left(\beta (1 + r_{t+z}) \left(\frac{U(c_{t+z})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho \\ Y_t = \delta + Y_{t-z} + \epsilon_t \quad \sqrt{n} \frac{y_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0, 1)$$

$$\omega = \eta(R\beta - r)' \left[R \hat{\Sigma}_{XX} \hat{S} \hat{\Sigma}_{XX}^{-1} R' \right]^{-1} (R\beta - r) \xrightarrow{d} \chi_m^2 \quad \hat{\zeta}(e) = \frac{1}{\tau_h} \sum_{t=1}^{\tau} K \left(\frac{\hat{e}_t - e}{h} \right) \quad l(\lambda; y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i!) \\ S^{AW} = \hat{\Gamma}_\theta + \sum_{i=1}^l \frac{l+1-i}{l+1} \left(\hat{\Gamma}_i + \hat{\Gamma}'_i \right) \quad Y_i = \beta_1 x_i + \beta_2 x_i I_{[x_i > \kappa]} + \epsilon_i \quad \sqrt{\tau} (R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right)$$

$$\rho_s = \frac{\gamma_s}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])]}{V[y_t]} \Rightarrow -2X'y - 2X'X\beta = \sigma \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y,x} \\ \text{Power}(\mu_1, \mu_2, \sigma, \alpha, N) = \Phi \left(\frac{\mu_2 - \mu_1}{\sqrt{\frac{\sigma}{N}}} \right) + \alpha - \Phi \left(\frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma}{N}}} \right)$$

$$\frac{\partial l(\theta; y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2} \quad A \cap B = \{x : x \in A \text{ and } x \in B\} \quad \sigma^2 = \omega + \alpha V_{t-1}^2 + \beta \sigma_{t-1}^2 \\ \sigma^2 = CC' + AA' \odot \epsilon_{t-1} \epsilon_{t-1}' + BB' \odot \Sigma_{t-1}$$

$$\text{VaR}_{t+z} = -\mu - \sigma_{t+z} G_{C^2}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial l(y; \psi)}{\partial \psi} \frac{\partial l(y; \psi)}{\partial \psi'} \right] \quad \hat{I}^1(x_1 | x_2 \in B) = \frac{\int_B \hat{\ell}^1(x_1, x_2) dx_2}{\int_B \hat{\ell}^1(x_2) dx_2} \\ \lambda_{\text{trace}}(r) = -\tau \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$z_t = \Upsilon z_{t-1} + \xi_t \quad \sigma^2 = \omega + \alpha V_{t-1}^2 + \beta \sigma_{t-1}^2 \\ \sigma^2 = CC' + AA' \odot \epsilon_{t-1} \epsilon_{t-1}' + BB' \odot \Sigma_{t-1}$$

Percent Value-at-Risk

- Value-at-Risk can be normalized and reported as a %

Definition (Percentage Value-at-Risk)

The α percentage Value-at-Risk ($\%VaR$) of a portfolio is defined as the largest return such that the probability that the return on the portfolio over some period of time is less than $-\%VaR$ is α ,

$$Pr(r < -\%VaR) = \alpha$$

where r is the percentage return on the portfolio. $\%VaR$ can be equivalently defined as $\%VaR = VaR/W_0$.

- Units are returns (no units)
- Also almost always *positive*
- Lets VaR be interpreted without knowing the value of the portfolio, W_0
 - ▶ No meaningful loss of information from standard VaR
 - ▶ Used throughout rest of lecture in place of formal definition of VaR

Relationship between Quantiles and VaR

- VaR is a quantile
 - ▶ Quantile of the future distribution

Definition (α -Quantile)

The α -quantile of a random variable X is defined as the *smallest* number q_α such that

$$\Pr(X \leq q_\alpha) = \alpha$$

- Other “-iles”
 - ▶ Tercile
 - ▶ Quartile
 - ▶ Quintile
 - ▶ Decile
 - ▶ Percentile

Conditional and Unconditional VaR

- Conditional VaR is similar to conditional mean or conditional variance

Definition (Conditional Value-at-Risk)

The conditional α Value-at-Risk is defined as

$$\Pr(r_{t+1} < -VaR_{t+1|t} | \mathcal{F}_t) = \alpha$$

where $r_{t+1} = (W_{t+1} - W_t) / W_t$ is the return on a portfolio at time $t + 1$.

- t is an arbitrary measure of time $\Rightarrow t + 1$ also refers to an arbitrary unit of time
 - ▶ day, two-weeks, 5 years, etc.
- Incorporates all information available at time t to assess risk at time $t + 1$
- Natural extension of conditional expectation and conditional variance to **conditional quantile**

Review

Key Concepts

Value-at-Risk, Quantile, Percentile, Conditional Value-at-Risk

Questions

- What are the units of Value-at-Risk?
- Is VaR normally positive or negative?
- Why is percent VaR simpler to use than VaR?
- How is 95%-VaR related to a quantile?

Problems

1. Suppose $W_0 = 2,500,000,000$ and $r_1 \sim N(0.15/12, 0.3^2/12)$. What are the 95% and 99% VaR?
2. What are the percent VaR in the previous problem?
3. What are the 95% and 99% VaRs if $r_1 \sim \text{Standardized } t_6$ with the same mean and variance?

Conditional VaR: RiskMetrics

- Industry standard benchmark
- Restricted GARCH(1,1)

$$\sigma_{t+1}^2 = (1 - \lambda)r_t^2 + \lambda\sigma_t^2$$

- *Exponentially Weighted Moving Average (EWMA) Variance:*

$$\sigma_{t+1}^2 = \sum_{i=0}^{\infty} (1 - \lambda)\lambda^i r_{t-i}^2$$

$$VaR_{t+1} = -\sigma_{t+1}\Phi^{-1}(\alpha)$$

- $\Phi^{-1}(\cdot)$ is the inverse normal cdf
- Advantages
 - ▶ No parameters to estimate
 - ▶ $\lambda = .94$ (daily data), $.97$ (weekly), $.99$ (monthly)
 - ▶ Easy to extend to portfolios (see notes)
- Disadvantages
 - ▶ No parameters to estimate
 - ▶ No leverage effect
 - ▶ Random Walk VaR

Conditional VaR: GARCH models for Value-at-Risk

$$r_{t+1} = \mu + \epsilon_{t+1}$$

$$\sigma_{t+1}^2 = \omega + \gamma \epsilon_t^2 + \beta \sigma_t^2$$

$$\epsilon_{t+1} = \sigma_{t+1} e_{t+1}$$

$$e_{t+1} \stackrel{\text{i.i.d.}}{\sim} f(0, 1)$$

- Value-at-Risk:

$$VaR_{t+1} = -\hat{\mu} - \hat{\sigma}_{t+1} F_\alpha^{-1}$$

- F_α^{-1} is the α quantile of the distribution of e_{t+1}
 - ▶ For example, 1.645 for the 5% from a normal

- Advantages
 - ▶ Flexible volatility model and easy to estimate

- Disadvantages
 - ▶ Must chose f (know f to get the correct VaR)
 - ▶ Location-Scale families

Conditional VaR: Semiparametric/Filtered HS

- Parametric GARCH + Nonparametric Density \rightarrow Semi-parametric VaR

$$e_{t+1} \stackrel{\text{i.i.d.}}{\sim} g(0, 1), \quad g \text{ unknown distribution}$$

- Implementation

1. Fit an ARCH model using Normal QMLE
2. $\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$
3. Order residuals

$$\hat{e}_1 < \hat{e}_2 < \dots < \hat{e}_{N-1} < \hat{e}_N$$

- Quantile is residual $\alpha \times N$ residual (N=T).

$$VaR_{t+1}(\alpha) = -\hat{\mu} - \hat{\sigma}_{t+1} \hat{G}_\alpha^{-1}$$

- Advantages

- ▶ All advantages of GARCH
- ▶ Quantile converges to true quantile

- Disadvantages

- ▶ Location-Scale families
- ▶ Quantile convergence is slow

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Conditional VaR: CaViaR

VaR, ES and Density Forecasting

$$\Delta y_t = \phi_\theta + \delta_{\pi} t + \gamma y_{t-\pi} + \sum_{\rho=1}^p \phi_\rho \Delta y_{t-\rho} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \\ t = \frac{\sqrt{n}(\hat{R}\hat{\theta} - r)}{\sqrt{RG^{-1}\Sigma(G^{-1})'R'}} \xrightarrow{d} N(\mu, 1) \quad BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\mathcal{Z}^4] \quad N(\mu_1 + \beta'(x_2 - \mu_2), \Sigma_{22} - \beta'\Sigma_{21}\beta)$$

$$\hat{\ell}^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \text{argmin}_{\beta} (y - X\beta)(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j| \\ \hat{\ell}^1(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{1-\pi}(1 - \rho)^{\beta-\pi}}{B(\alpha, \beta)} \quad E \left[\left(\beta(1 + r_{t+\pi}) \left(\frac{U(c_{t+1})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho$$

$$\left[\begin{array}{c} \Delta x_i \\ \Delta y_i \end{array} \right] = \left[\begin{array}{c} \pi_{x2} \\ \pi_{y2} \end{array} \right] + \left[\begin{array}{c} \alpha_x e_i \\ \alpha_y e_i \end{array} \right] + \pi_{\pi} \left[\begin{array}{c} \Delta x_{t-\pi} \\ \Delta y_{t-\pi} \end{array} \right] + \dots + \pi_{\pi} \left[\begin{array}{c} \Delta x_{2\pi} \\ \Delta y_{2\pi} \end{array} \right] + \left[\begin{array}{c} \eta_{x,i} \\ \eta_{y,i} \end{array} \right]$$

$$\omega = \eta(R\beta - r)' \left[R \hat{\Sigma}_{XX} \hat{S} \hat{\Sigma}_{XX}^{-1} R' \right]^{-1} (R\beta - r) \xrightarrow{d} \chi_m^2 \quad Y_4 = \delta + Y_{t-\pi} + \epsilon_t \\ \mathcal{G}(\epsilon) = \frac{1}{\tau_h} \sum_{t=2}^{\tau} K \left(\frac{\hat{\epsilon}_t - \epsilon}{h} \right) \quad f(\lambda, y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i)$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad S^{\Delta\omega} = \tilde{\Gamma}_\theta + \sum_{i=1}^l \frac{l+\pi-\zeta}{l+\pi} \left(\tilde{\Gamma}_i + \tilde{\Gamma}'_i \right) \quad Y_\zeta = \beta_{12} X_\zeta + \beta_{22} X_\zeta I_{[X_\zeta > \kappa]} + \epsilon_\zeta \\ \sqrt{\tau}(R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right)$$

$$\rho_\pi = \frac{\gamma_\pi}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-\pi} - E[y_{t-\pi}])]}{V[y_t]} \quad -2X'y = -2X'\hat{\epsilon} = \nu \quad \beta \approx \frac{\partial Y_\zeta}{\partial X_\zeta} \frac{X_\zeta}{Y_\zeta} = E_{y, x} \\ \beta = \frac{\partial Y_\zeta}{\partial X_\zeta} \frac{X_\zeta}{Y_\zeta} = E_{y, x} \quad \frac{\partial l(\theta, y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2} \quad \mathbb{A} \cap \mathbb{B} = \{x : x \in \mathbb{A} \text{ and } x \in \mathbb{B}\}$$

$$\rho_{\pi\pi} = \frac{\gamma_{\pi\pi}}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-\pi} - E[y_{t-\pi}])]}{V[y_t]} \quad -2X'y + 2X'X\beta = \nu \quad \beta \approx \frac{\partial Y_\zeta}{\partial X_\zeta} \frac{X_\zeta}{Y_\zeta} = E_{y, x} \\ \frac{\partial l(\theta, y)}{\partial \sigma^2} = -\frac{\nu}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^4} \quad \lambda_{trace}(r) = -\tau \sum_{i=r+\pi}^{\infty} \ln(1 - \hat{\lambda}_i) \quad f(x_1 | X_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2} \\ f(x_1 | X_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$\text{VaR}_{t+\pi} = -\mu - \sigma_{t+\pi} G_{C^2}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial l(y; \psi)}{\partial \psi} \frac{\partial l(y; \psi)}{\partial \psi'} \right] \quad \mathcal{C}(u_1, u_2, \dots, u_k) = \frac{\partial^k C(u_1, u_2, \dots, u_k)}{\partial u_1 \partial u_2 \dots \partial u_k}$$

$$\sigma^2 = \omega + \alpha V_{t-\pi}^2 + \beta \sigma_{t-\pi}^2 \quad \Sigma_t = CC' + AA' \odot \epsilon_{t-\pi} \epsilon_{t-\pi} + BB' \odot \Sigma_{t-\pi} \\ \sigma^2 = \omega + \alpha V_{t-\pi}^2 + \beta \sigma_{t-\pi}^2$$

$$\lambda_{trace}(r) = -\tau \sum_{i=r+\pi}^{\infty} \ln(1 - \hat{\lambda}_i) \quad f(x_1 | X_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2} \\ f(x_1 | X_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$\sigma^2 = \omega + \alpha V_{t-\pi}^2 + \beta \sigma_{t-\pi}^2 \quad \Sigma_t = CC' + AA' \odot \epsilon_{t-\pi} \epsilon_{t-\pi} + BB' \odot \Sigma_{t-\pi}$$

$$\sigma^2 = \omega + \alpha V_{t-\pi}^2 + \beta \sigma_{t-\pi}^2 \quad \Sigma_t = CC' + AA' \odot \epsilon_{t-\pi} \epsilon_{t-\pi} + BB' \odot \Sigma_{t-\pi}$$

Conditional VaR: CaViaR

- Conditional Autoregressive Value-at-Risk (ARCVaR)
 - ▶ Conditional quantile *regression*
 - ▶ Directly parameterize quantile $F_\alpha^{-1} = q$ of the return distribution

$$q_{t+1} = \omega + \gamma HIT_t + \beta q_t$$

$$HIT_t = I_{[r_t < q_t]} - \alpha$$

$$VaR_{t+1} = -q_{t+1}$$

- Advantages
 - ▶ Focuses on quantile
 - ▶ Flexible specification
- Disadvantages
 - ▶ Hard to estimate
 - ▶ Which specification?
 - ▶ Out-of-order VaR: 5% can less than 10% VaR

Estimation of CaViaR models

- Many CaViaR specifications

- ▶ Symmetric

$$q_{t+1} = \omega + \gamma HIT_t + \beta q_t.$$

- ▶ Symmetric absolute value,

$$q_{t+1} = \omega + \gamma |r_t| + \beta q_t.$$

- ▶ Asymmetric absolute value

$$q_{t+1} = \omega + \gamma_1 |r_t| + \gamma_2 |r_t| I_{[r_t < 0]} + \beta q_t$$

- ▶ Indirect GARCH

$$q_{t+1} = (\omega + \gamma r_t^2 + \beta q_t^2)^{\frac{1}{2}}$$

- Estimation minimizes the “tick” loss function

$$\operatorname{argmin}_{\theta} T^{-1} \sum_{t=1}^T \alpha(r_t - q_t)(1 - I_{[r_t < q_t]}) + (1 - \alpha)(q_t - r_t)I_{[r_t < q_t]}$$

- ▶ Non-differentiable

- ▶ Requires “derivative-free” optimizers (e.g. simplex optimizers)

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{z}{\sqrt{n}} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Weighted Historical Simulation

VaR, ES and Density Forecasting

$$\Delta y_t = \phi_\theta + \delta_z t + \gamma y_{t-z} + \sum_{p=1}^p \phi_p \Delta y_{t-p} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \\ BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\epsilon^4] \quad N(\mu_1 + \beta' (x_2 - \mu_2), \Sigma_{22} - \beta' \Sigma_{21} \beta) \\ f^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \arg\min_{\beta} (y - X\beta)(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j|$$

$$\left[\begin{array}{c} \Delta x_i \\ \Delta y_i \end{array} \right] = \left[\begin{array}{c} \pi_{x2} \\ \pi_{y2} \end{array} \right] + \left[\begin{array}{c} \alpha_x e_i \\ \alpha_y e_i \end{array} \right] + \pi_{\pi} \left[\begin{array}{c} \Delta x_{t-2} \\ \Delta y_{t-2} \end{array} \right] + \dots + \pi_{\pi} \left[\begin{array}{c} \Delta x_{t-k} \\ \Delta y_{t-k} \end{array} \right] + \left[\begin{array}{c} \eta_{t+k} \\ \eta_{t+k} \end{array} \right]$$

$$\hat{f}(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{n-p}(1 - \rho)^{\beta - p}}{B(\alpha, \beta)} \quad E \left[\left(\beta (1 + r_{t+p}) \left(\frac{U(c_{t+2})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho$$

$$\omega = \eta(R\beta - r)' \left[R \hat{\Sigma}_{XX} \hat{S} \hat{\Sigma}_{XX}^{-1} R' \right]^{-1} (R\beta - r) \xrightarrow{d} \chi_m^2 \quad Y_t = \delta + Y_{t-z} + \epsilon_t$$

$$g(e) = \frac{1}{\tau_h} \sum_{t=2}^{\tau} K \left(\frac{\hat{e}_t - e}{h} \right) \quad l(\lambda, y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i!)$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad S^{\text{EW}} = \bar{\Gamma}_\theta + \sum_{i=1}^l \frac{l + 1 - i}{l + 1} \left(\bar{\Gamma}_i + \bar{\Gamma}'_i \right) \quad \sqrt{\tau}(R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right)$$

$$\rho_s = \frac{\gamma_s}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])]}{V[y_t]} \Rightarrow -2X'y - 2X'X\beta = \sigma \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y,x}$$

$$\text{VaR}_{t+z} = -\mu - \sigma_{t+z} G_{C^2}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial J(y; \psi)}{\partial \psi} \frac{\partial J(y; \psi)}{\partial \psi'} \right]$$

$$\sum_{i=1}^n g_i(Y_i) \xrightarrow{d} E[g(Y)] \quad \frac{\partial l(\theta; y)}{\partial \sigma^2} = -\frac{\alpha}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^4}$$

$$A \cap B = \{x : x \in A \text{ and } x \in B\} \quad \frac{\partial l(\theta; y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2}$$

$$f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B \hat{f}_2(x_2) dx_2} \quad \lambda_{\text{trace}}(r) = -\tau \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i)$$

$$f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B \hat{f}_2(x_2) dx_2} \quad \sigma_t^2 = \omega + \alpha V_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$\Sigma_t = CC' + AA' \odot \epsilon_{t-1} \epsilon_{t-1}' + BB' \odot \Sigma_{t-1}$$

Weighted Historical Simulation

- Uses a weighted empirical cdf
- Weights are exponentially decaying

$$w_i = \lambda^{t-i} (1 - \lambda) / (1 - \lambda^t), \quad i = 1, 2, \dots, t$$

- Weighted Empirical cdf

$$\hat{G}_t(r) = \sum_{i=1}^t w_i I_{[r_i < r]}$$

- Conditional VaR is solution to

$$\text{VaR}_{t+1} = \min_r \hat{G}(r) \geq \alpha$$

- Example uses $\lambda = 0.975$

Review

Key Concepts

RiskMetrics, Filtered Historical Simulation, CaViaR, Weighted Historical Simulation, Historical Simulation

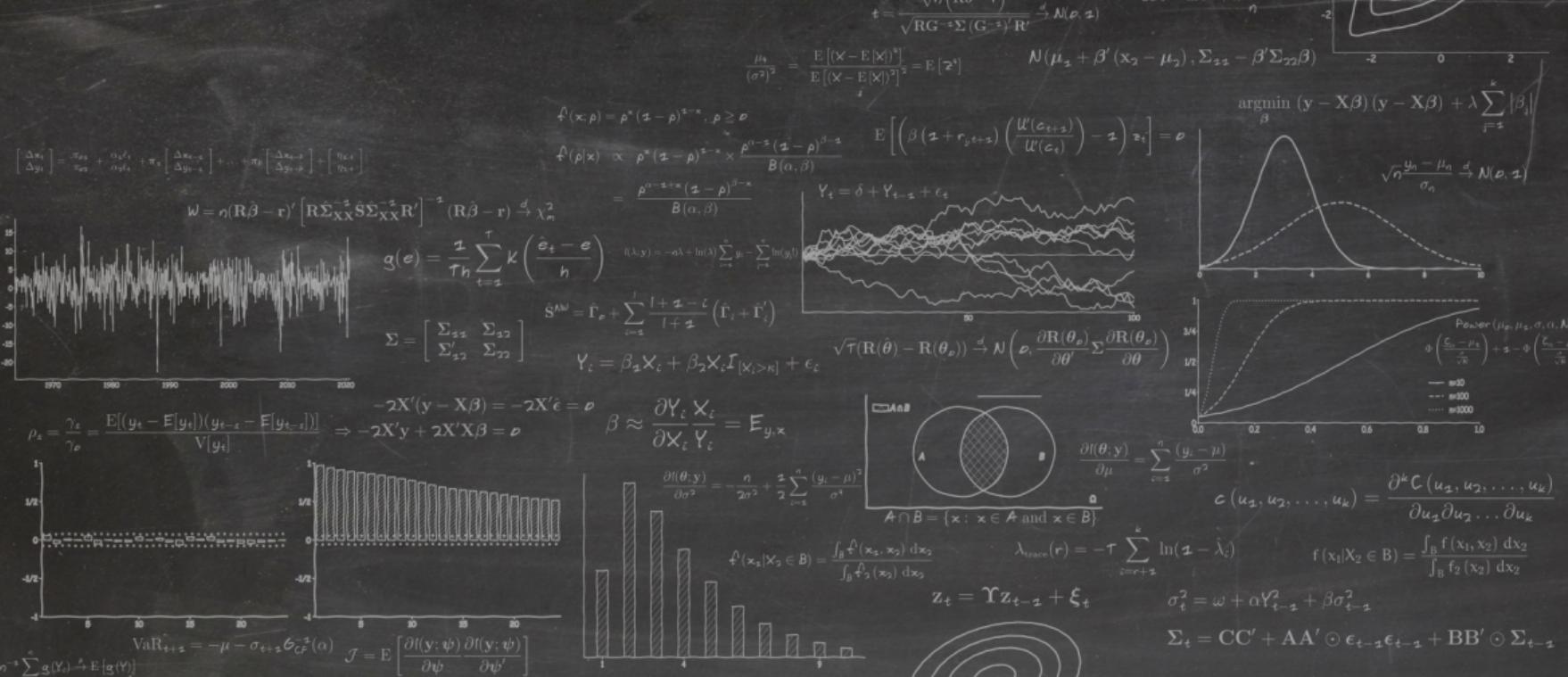
Questions

- How are filtered, weighted and standard historical simulation similar? How do they differ?
- What is the key limitation to historical simulation as a method to forecast VaR?
- When do CaViaR and GARCH produce identical VaR forecasts?
- What is the key challenge when fitting a CaViaR model?
- How does the CaViaR model differ from FHS?

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Evaluation of Value-at-Risk Models

VaR, ES and Density Forecasting



Evaluating VaR models

- Basic instrument for testing VaR is the “Hit”

$$get = I_{[r_t < F_t^{-1}]} - \alpha = HIT_t$$

- Is the *generalized error* from the “tick” loss function
- If the VaR is correct,

$$\mathbb{E}_{t-1}[HIT_t] = 0$$

- Leads to a standard Generalized Mincer-Zarnowitz evaluation framework
- Hit Regression

$$HIT_{t+h} = \gamma_0 + \gamma_1 VaR_{t+h|t} + \gamma_2 HIT_t + \gamma_3 HIT_{t-1} + \dots + \gamma_K HIT_{t-K+1}$$

- ▶ Null is $H_0 : \gamma_0 = \gamma_1 = \dots = \gamma_K = 0$
- ▶ Alternative is $H_1 : \gamma_j \neq 0$ for some j
- As always, GMZ can be augmented with any time t measurable variable

Unconditional Evaluation of VaR using the Bernoulli

- \widetilde{HIT} s from a correct VaR model have a Bernoulli distribution
 - ▶ 1 with probability α
 - ▶ 0 with probability $1 - \alpha$
- Likelihood for T Bernoulli random variables $x_t \in \{0, 1\}$

$$f(x_t; p) = \prod_{t=1}^T p^{x_t} (1-p)^{1-x_t}$$

- Log-likelihood is

$$l(p; x_t) = \sum_{t=1}^T x_t \ln p + (1 - x_t) \ln 1 - p$$

- In terms of α and \widetilde{HIT}_t

$$l(\alpha; \widetilde{HIT}_t) = \sum_{t=1}^T \widetilde{HIT}_t \ln \alpha + (1 - \widetilde{HIT}_t) \ln 1 - \alpha$$

- Easy to conduct a LR test

$$LR = 2(l(\hat{\alpha}; \widetilde{HIT}) - l(\alpha_0; \widetilde{HIT})) \sim \chi_1^2$$

- $\hat{\alpha} = T^{-1} \sum_{t=1}^T \widetilde{HIT}_t$, α_0 is the α from the VaR

Evaluation of Conditional VaR using the Bernoulli

- Can also be extended to testing conditional independence of *HITs*
- Define

$$\begin{aligned} n_{00} &= \sum_{t=1}^{T-1} (1 - \widetilde{\text{HIT}}_t)(1 - \widetilde{\text{HIT}}_{t+1}), & n_{10} &= \sum_{t=1}^{T-1} (1 - \widetilde{\text{HIT}}_t) \widetilde{\text{HIT}}_{t+1} \\ n_{01} &= \sum_{t=1}^{T-1} \widetilde{\text{HIT}}_t (1 - \widetilde{\text{HIT}}_{t+1}), & n_{11} &= \sum_{t=1}^{T-1} \widetilde{\text{HIT}}_t \widetilde{\text{HIT}}_{t+1} \end{aligned}$$

- The log-likelihood for the sequence two VaR exceedences is

$$l(p; \widetilde{\text{HIT}}) = n_{11} \ln(p_{11}) + n_{01} \ln(1 - p_{11}) + n_{00} \ln(p_{00}) + n_{10} \ln(1 - p_{00})$$

Evaluation of Conditional VaR using the Bernoulli

- Null is $H_0 : p_{11} = 1 - p_{00} = \alpha$
- MLEs are

$$\hat{p}_{00} = \frac{n_{00}}{n_{00} + n_{10}}, \quad \hat{p}_{11} = \frac{n_{11}}{n_{11} + n_{01}}$$

- Tested using a likelihood ratio test

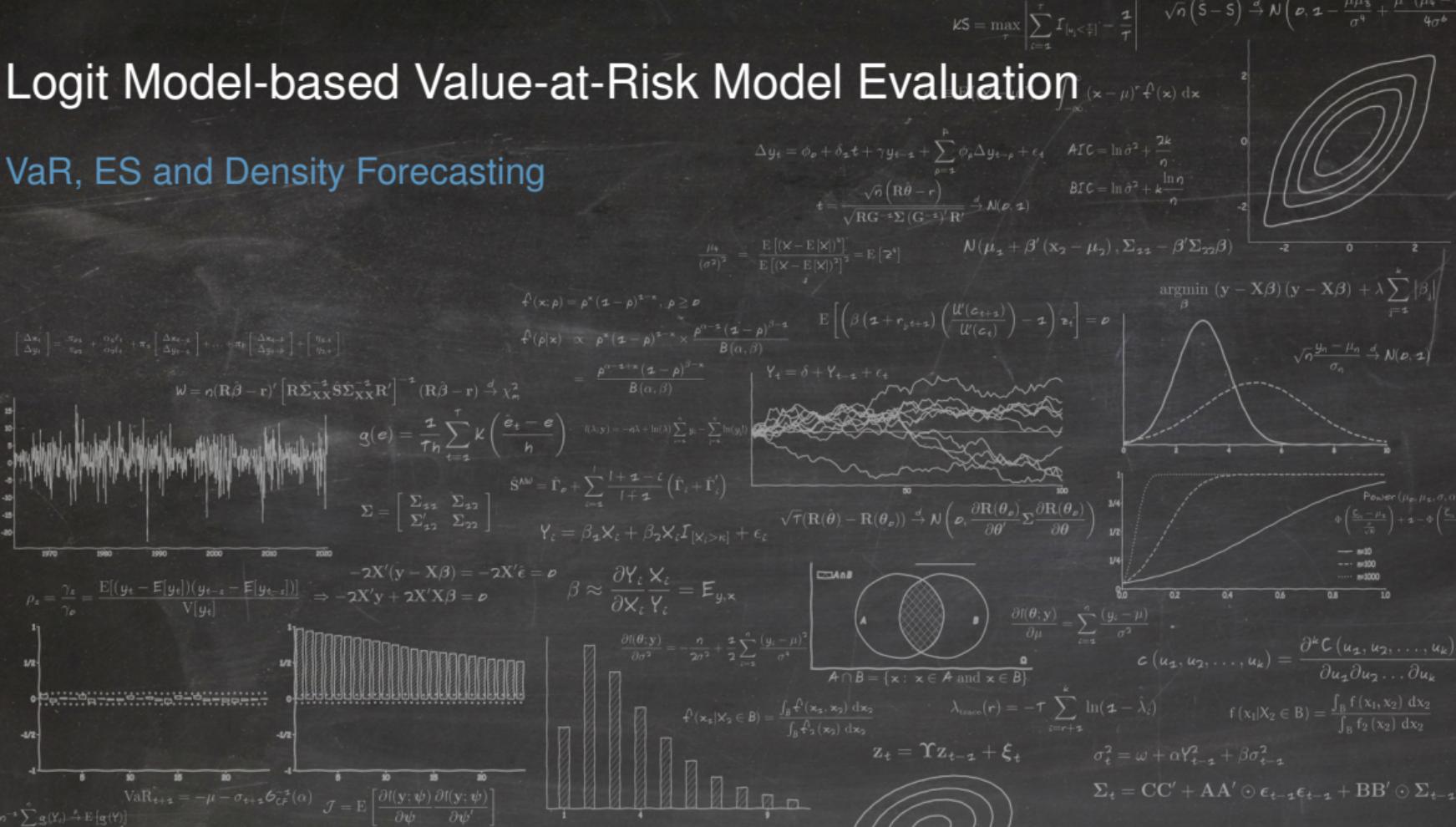
$$LR = 2(l(\hat{p}_{00}, \hat{p}_{11}; \widetilde{HIT}) - l(p_{00} = 1 - \alpha, p_{11} = \alpha; \widetilde{HIT}))$$

- Test statistic follows a χ^2_2 distribution

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Logit Model-based Value-at-Risk Model Evaluation

VaR, ES and Density Forecasting



Relationship to Probit/Logit

- Standard GMZ regression is not an ideal test
- Ignores special structure of a *HIT*
- A *HIT* is a limited dependent variable
 - ▶ Only takes one of two values
- Define a modified hit $\widetilde{HIT}_t = I_{[r_t < F_t^{-1}]}$
 - ▶ Takes the value 1 with probability α and 0 with probability $1 - \alpha$
 - ▶ Name that distribution →
- Leads to a modified regression framework known as a probit or logit
 - ▶ Probit:
$$\widetilde{HIT}_{t+1} = \Phi(\gamma_0 + \mathbf{x}_t \boldsymbol{\gamma})$$
 - If model is correct, $\gamma_0 = \Phi^{-1}(\alpha)$ and $\boldsymbol{\gamma} = \mathbf{0}$
 - Estimated using Bernoulli Maximum Likelihood
 - Easy to compute Likelihood ratio
- Accounts for the limited range of the variable and that the density is non-normal
- Allows for simple-yet-powerful likelihood ratio tests under the null

Review

Key Concepts

Generalized Error, Tick loss function, HIT, Christoffersen's Test, Probit, Logit

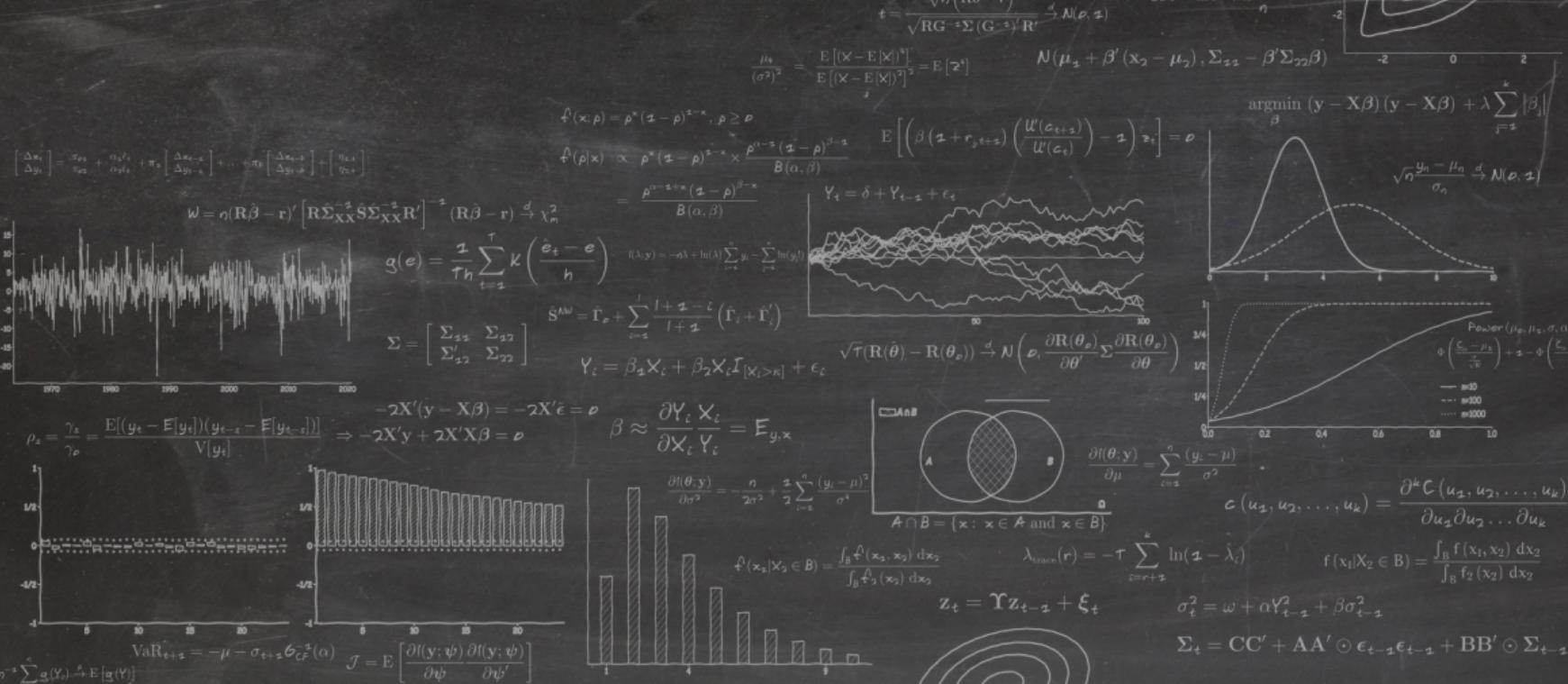
Questions

- How is VaR model evaluation similar to mean and volatility model evaluation?
- What special features for VaR forecasts distinguish them from mean and volatility forecasts?
- Why should likelihood methods outperform OLS when evaluating VaR models?
- What is the key limitation of the Christoffersen's test?

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Density Forecasting and Evaluation

VaR, ES and Density Forecasting



Density Estimation and Forecasting

- End all be all of risk measurement
- Issues:
 - ▶ Equally hard
 - ▶ Lots of estimation and model error
 - Can have non obvious effects on nonlinear functions (i.e. options)
 - ▶ Not closed under aggregation
 - No multi-step
- Builds off of the GARCH VaR application

Density forecasts from GARCH models

- Simple constant mean GARCH(1,1)

$$r_{t+1} = \mu + \epsilon_{t+1}$$

$$\sigma_{t+1}^2 = \omega + \gamma \epsilon_t^2 + \beta \sigma_t^2$$

$$\epsilon_{t+1} = \sigma_{t+1} e_{t+1}$$

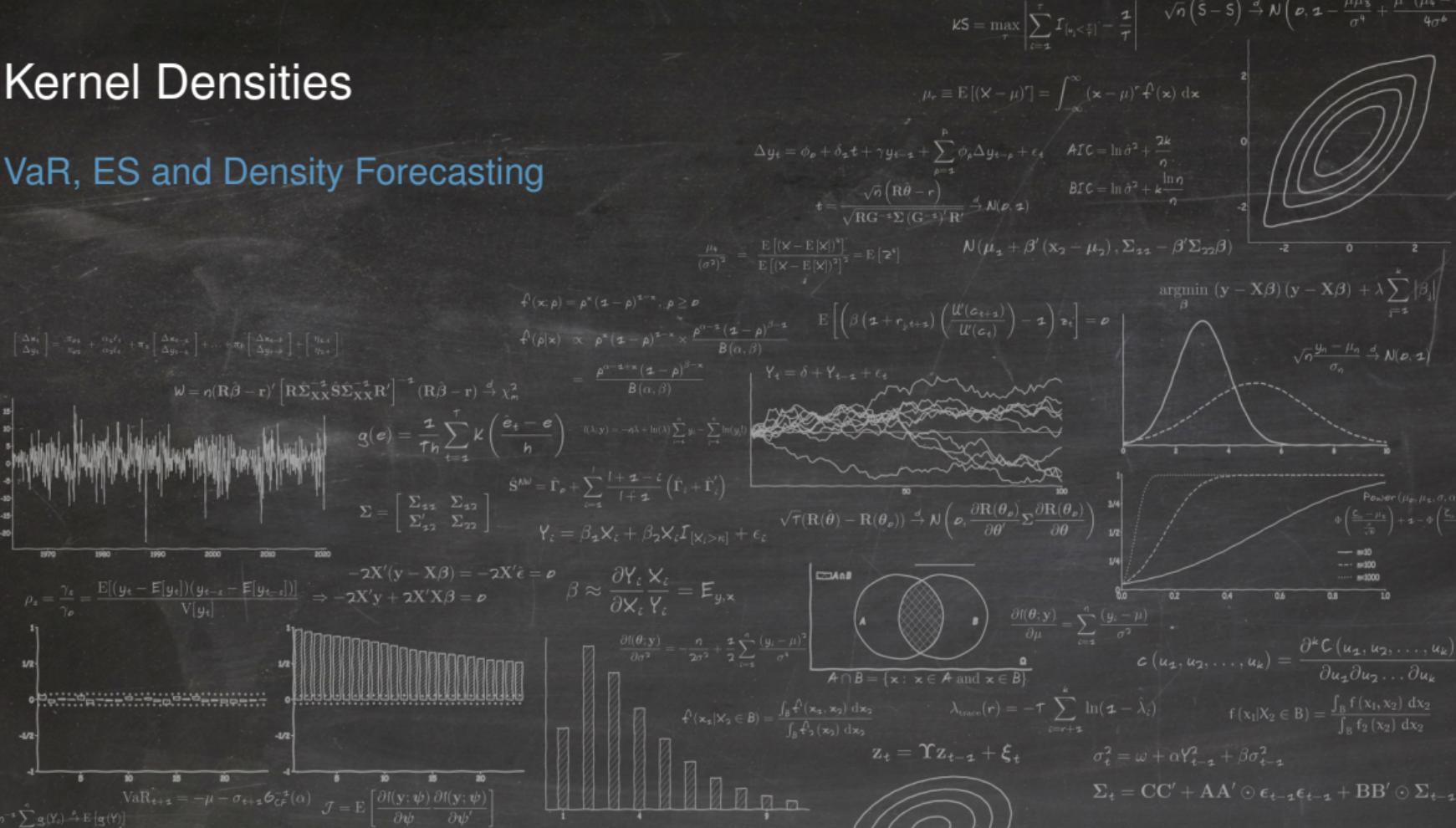
$$e_{t+1} \stackrel{\text{i.i.d.}}{\sim} g(0, 1).$$

- g is some known distribution, but not necessarily normal
- Density forecast is simply $g(\mu, \sigma_{t+1|t}^2)$
- Flexible through choice of g
- Parsimonious
- Semiparametric works in same way replacing g with the standardized residuals of a “smoothed” estimate

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Kernel Densities

VaR, ES and Density Forecasting



Kernel Densities

- “Smoothed” densities are more precise than rough estimates

$$g(e) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{\hat{e}_t - e}{h}\right), \quad \hat{e}_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t} = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

- Local average of how many \hat{e}_t there are in a small neighborhood of e
- $K(\cdot)$ is a kernel
 - ▶ Gaussian
 - ▶ Epanechnikov

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- h : Bandwidth controls smoothing
- Silverman’s bandwidth

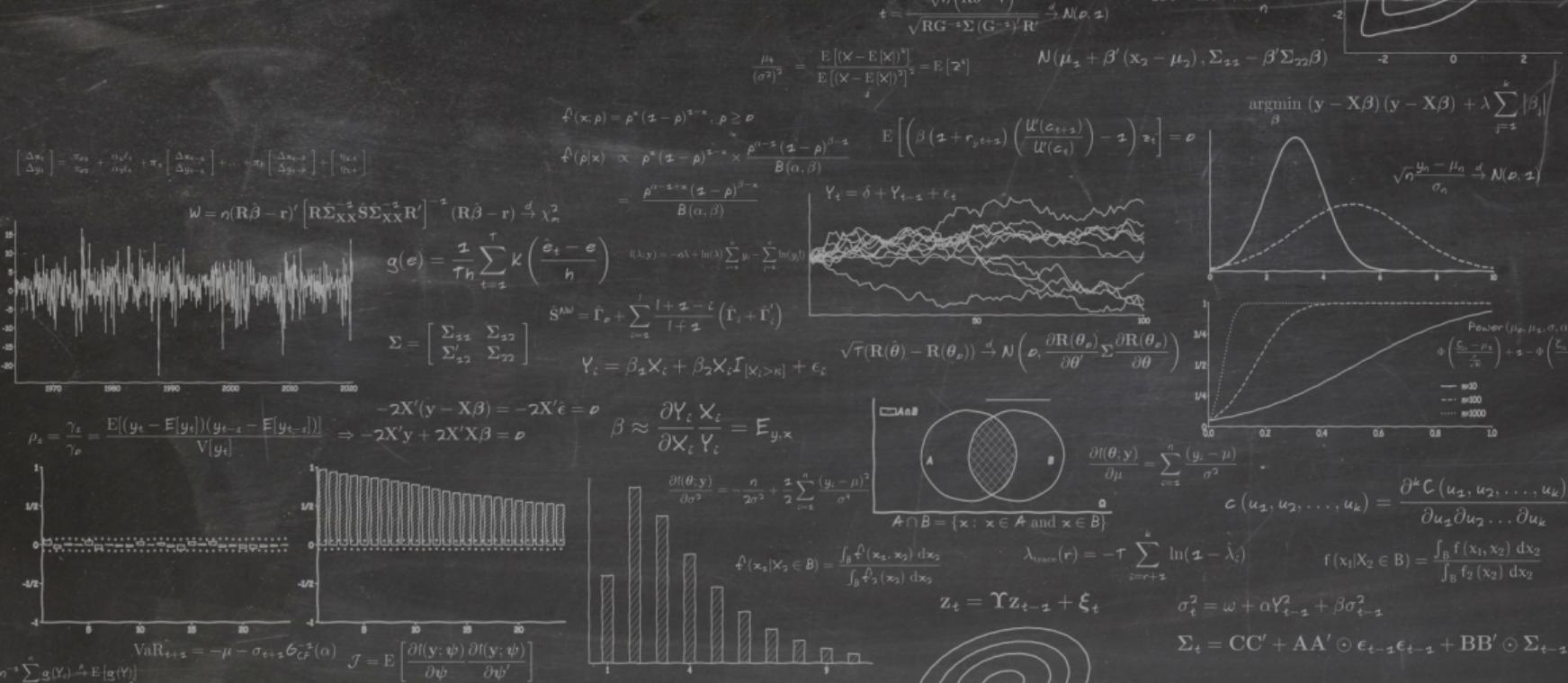
$$1.06\sigma_x T^{-\frac{1}{5}}$$

- ▶ h too small produces very rough densities (low bias but lots of variance)
- ▶ h too large produces overly smooth (low variance but very biased)

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Multi-step Density Forecasting

VaR, ES and Density Forecasting



Multi-step Density Forecasts

- Densities do not aggregate in general
 - ▶ Multivariate normal is special
- Densities from GARCH models do not easily aggregate
- 1-step density forecast from a standard GARCH(1,1)

$$r_{t+1} | \mathcal{F}_t \sim N(\mu, \sigma_{t+1|t}^2)$$

- Wrong 2-step forecast from a standard GARCH(1,1)

$$r_{t+2} | \mathcal{F}_t \sim N(\mu, \sigma_{t+2|t}^2)$$

- Correct 2-step forecast from a standard GARCH(1,1)

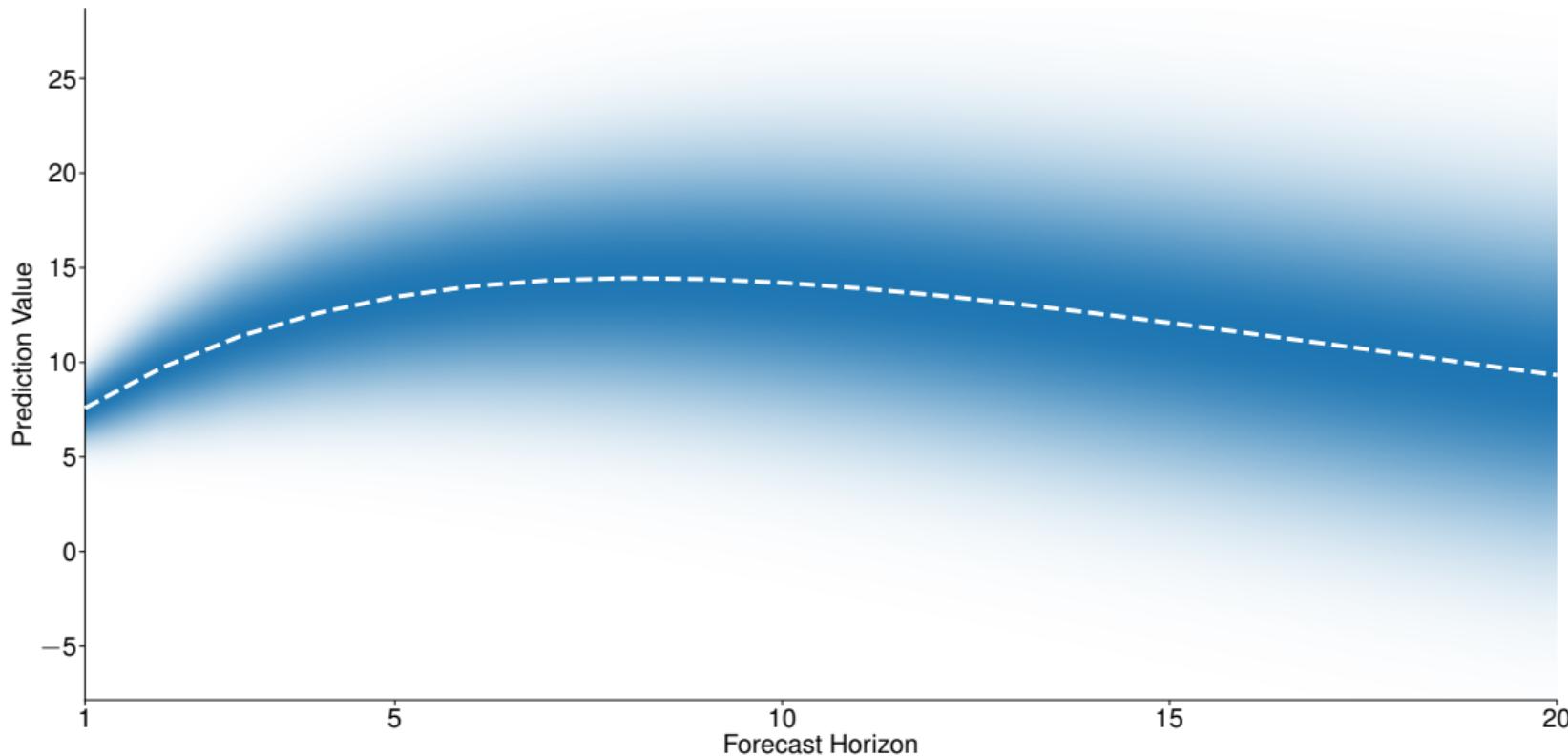
$$r_{t+2} | \mathcal{F}_t \sim \int_{-\infty}^{\infty} \phi(\mu, \sigma^2(e_{t+1})_{t+2|t+1}) \phi(e_{t+1}) de_{t+1}.$$

- Must integrate out the variance uncertainty between $t + 1$ and $t + 2$
- Easy fix: directly model $t + 2$ (or $t + h$)

The Fan plot

- Hard to produce time-series of densities
- Solution is the Fan Plot
- Popularized by the Bank of England
- Horizontal axis (x) is the number of time-periods ahead
- Vertical axis (y) is the value the variable might take
- Density is expressed using varying degrees of color intensity.
 - ▶ Dark colors indicate the highest probability
 - ▶ Progressively lighter colors represent decreasing likelihood
 - ▶ Essentially a plot of many quantiles of the distribution through time

A fan plot for an AR(2)



Review

Key Concepts

Density Forecast, Kernel Density, Bandwidth, Silverman's Bandwidth, Fan Plot

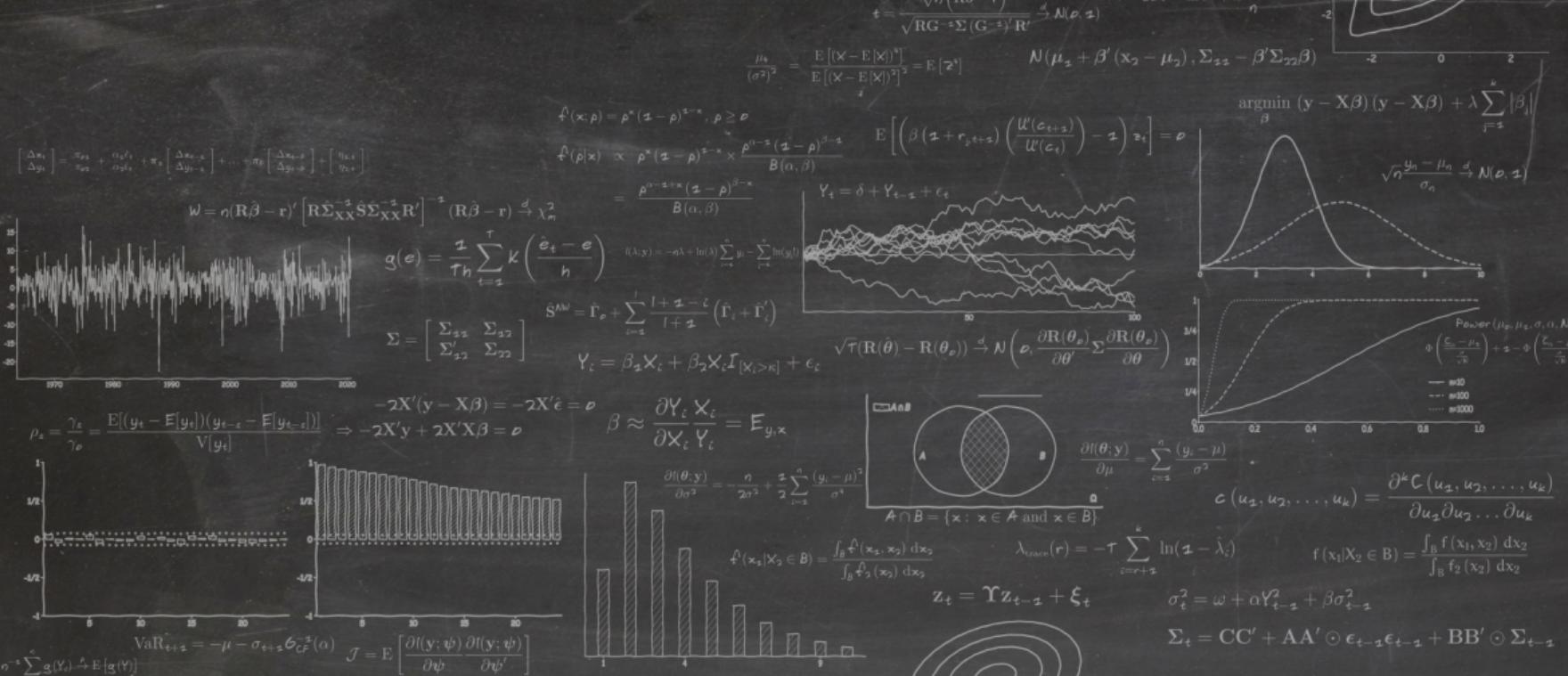
Questions

- What is the difference between a parametric and semi-parametric density forecast?
- How is a kernel density similar to a histogram?
- What does the bandwidth parameter in a kernel density determine?
- What do density forecasts not aggregate (in general)?

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{z}{\sqrt{n}} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

The Probability Integral Transformation

VaR, ES and Density Forecasting



Density “Standardized” Residuals

- Consider a generic stochastic process $\{y_t\}$

- ▶ Residuals from mean models:

$$\hat{\epsilon}_t = y_t - \hat{\mu}_t$$

- ▶ Residuals from variance models:

$$\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t} = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

- ▶ Residuals from Value-at-Risk models:

$$HIT_t = I_{[y_t < q_t]} - \alpha$$

- ▶ Residual from density models:

$$\hat{u}_t = F_t(y_t)$$

- Known as the Probability Integral Transformed Residuals

- One very useful property: If $y_t \sim F$ then $u_t \equiv F(y_t) \sim U(0, 1)$

Probability Integral Transform

Theorem (Probability Integral Transform)

Let a random variable X have a continuous, increasing cdf $F_X(x)$ and define $Y = F_X(X)$. Then Y is uniformly distributed and $\Pr(Y \leq y) = y$, $0 < y < 1$.

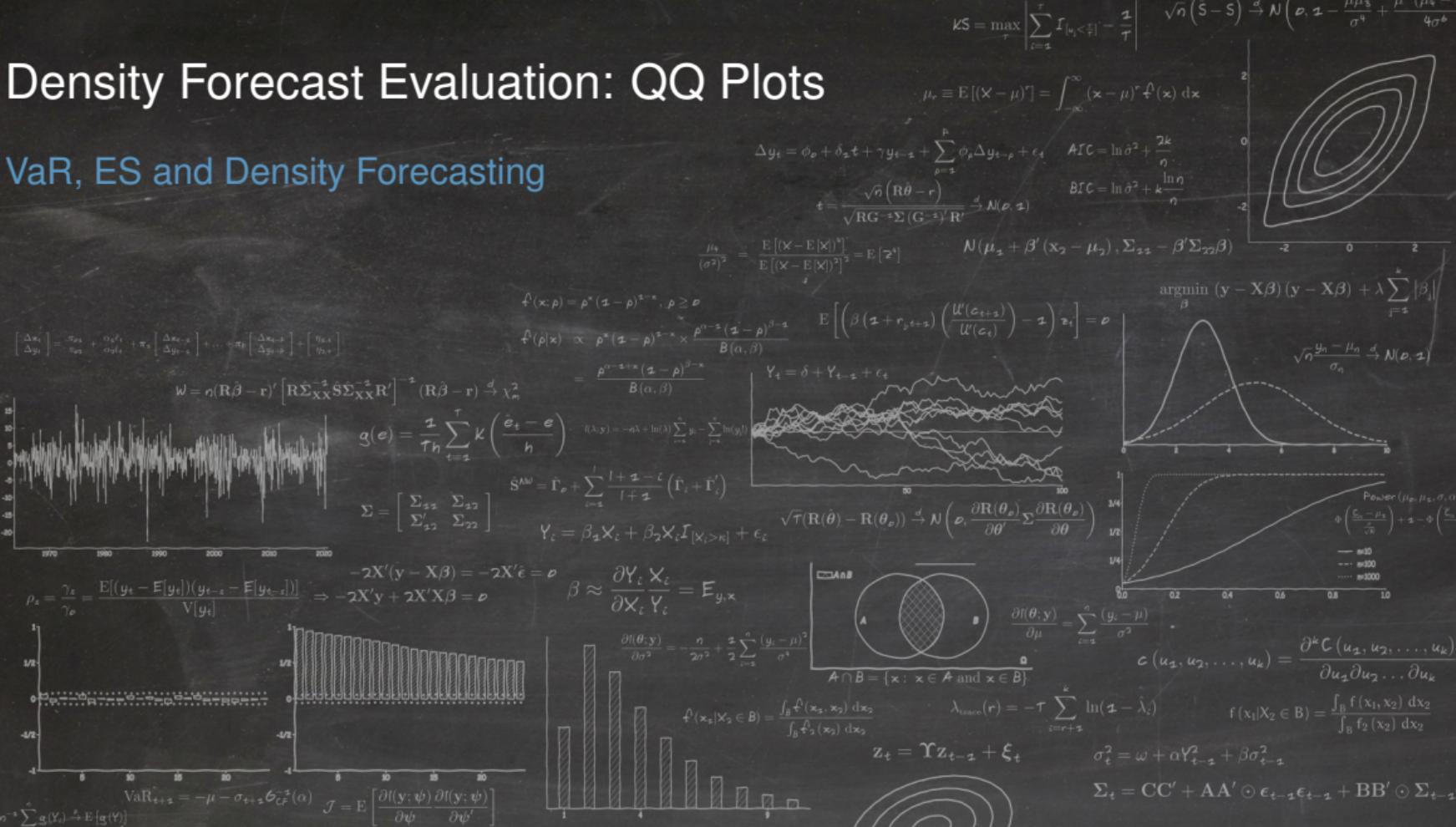
For any $y \in (0, 1)$, $Y = F_X(X)$, and so

$$\begin{aligned}\Pr(Y \leq y) &= \Pr(F_X(X) \leq y) \\ &= \Pr(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) && \text{Since } F_X^{-1} \text{ is increasing} \\ &= \Pr(X \leq F_X^{-1}(y)) && \text{Invertible since strictly increasing} \\ &= F_X(F_X^{-1}(y)) && \text{Definition of } F_X \\ &= y\end{aligned}$$

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Density Forecast Evaluation: QQ Plots

VaR, ES and Density Forecasting



Evaluating Density Forecasts: QQ Plots

- Quantile-Quantile Plots
- Plots the data against a hypothetical distribution

$$\hat{e}_1 < \hat{e}_2 < \dots < \hat{e}_{N-1} < \hat{e}_N$$

- ▶ $N = T$ but used to indicate that the index is not related to time
- e_n against $F^{-1}\left(\frac{j}{T+1}\right)$

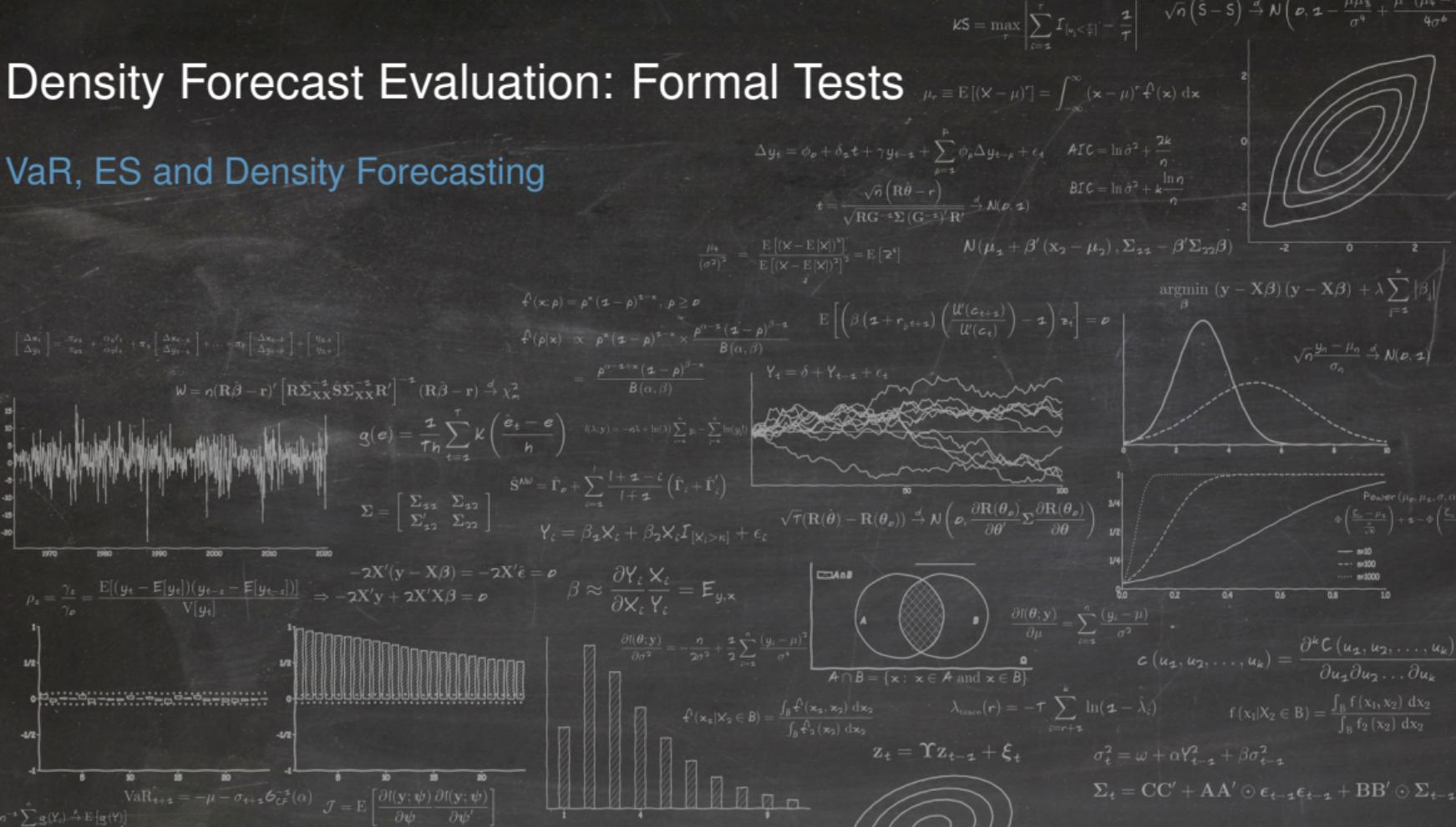
$$F^{-1}\left(\frac{1}{T+1}\right) < F^{-1}\left(\frac{2}{T+1}\right) < \dots < F^{-1}\left(\frac{T-1}{T+1}\right) < F^{-1}\left(\frac{T}{T+1}\right)$$

- F^{-1} is inverse cdf of distribution being used for comparison
- Should lie along a 45° line
- No confidence bands

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Density Forecast Evaluation: Formal Tests

VaR, ES and Density Forecasting



Evaluating Density Forecasts: Kolmogorov-Smirnov

- Formalizes QQ plots
- Key property
 - ▶ If $x \sim F$, then $u \equiv F(x) \sim U(0, 1)$
 - ▶ Can test $U(0, 1)$
- KS tests maximum deviation from $U(0, 1)$

$$\max_{\tau} \left| \frac{1}{T} \left(\sum_{i=1}^{\tau} I_{[u_i < \frac{\tau}{T}]} \right) - \frac{\tau}{T} \right|, \quad \tau = 1, 2, \dots, T$$

- ▶ $\frac{1}{T} \sum_{i=1}^{\tau} I_{[u_i < \frac{\tau}{T}]}$: Empirical percentage of u below τ/T
- ▶ τ/T : How many *should* be below τ/T
- Nonstandard distribution
- Parameter estimation error
 - ▶ Parameter Estimation Error (PEE) causes significant size distortions
 - ▶ Using a 5% CV will only reject 0.1% of the time
 - ▶ Solution is to simulate the needed critical values

Addressing PEE in a KS test

- Model is a complete model so can be easily simulated
- Exact KS distribution tabulated

Algorithm (Correct CV for KS test with PEE)

1. *Estimate model and save $\hat{\theta}$*
2. *Repeat many times (1000+)*
 - a. *Simulate artificial series from model using $\hat{\theta}$ with same number of observations as original data*
 - b. *Estimate parameters from simulated data, $\ddot{\theta}$*
 - c. *Compute KS test statistic on simulated data using $\ddot{\theta}$ and save as KS_i , $i = 1, 2, \dots$,*
3. *Sort the KS_i values and use the $1 - \alpha$ quantile for get correct CV for α size test*

Evaluating Density Forecasts: Berkowitz Test

- Berkowitz Test extends KS to evaluation of conditional densities
- Exploits probability integral transform property

$$\hat{u}_t = F(y_t)$$

- But then *re-transforms* the data to a standard normal

$$\hat{\eta}_t = \Phi^{-1}(\hat{u}_t) = \Phi^{-1}(F(y_t))$$

- ▶ Since $\hat{u}_t \stackrel{\text{i.i.d.}}{\sim} U(0, 1)$, $\hat{\eta}_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$
- Test is a likelihood ratio test using an AR(1)

$$\hat{\eta}_t = \phi_0 + \phi_1 \hat{\eta}_{t-1} + \nu_t$$

- Correctly specified: $H_0 : \phi_0 = 0 \cap \phi_1 = 0 \cap \sigma^2 = V[\nu_t] = 1$
- Likelihood ratio

$$2 \left(l(\eta_t | \hat{\phi}_0, \hat{\phi}_1, \hat{\sigma}^2) - l(\eta_t | \phi_0 = 0, \phi_1 = 0, \sigma^2 = 1) \right) \sim \chi^2_3$$

- ▶ Critical values wrong if F has estimated parameters

Review

Key Concepts

Probability Integral Transform, QQ Plot, Kolmogorov-Smirnov Test, Berkowitz Test

Questions

- How are PITs like forecast errors from ARMA models?
- What is plotted in a QQ plot?
- What does the KS test, and why should this quantity be small when the null is true?
- What is the consequence of ignoring PEE in a KS test?
- What does Berkowitz's test examine when evaluating a density?

$$KS = \max_{\tau} \left| \sum_{t=\pm}^{\tau} I_{[u_t < \frac{z}{\sqrt{n}}]} - \frac{1}{\tau} \right| \quad \sqrt{n} \left(\hat{S} - S \right) \xrightarrow{d} N \left(0, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

Expected Shortfall

VaR, ES and Density Forecasting

$$\Delta y_t = \phi_\theta + \delta_{xz} t + \gamma y_{t-z} + \sum_{\rho=1}^p \phi_\rho \Delta y_{t-\rho} + \epsilon_t \quad AIC = \ln \hat{\sigma}^2 + \frac{2k}{n} \\ BIC = \ln \hat{\sigma}^2 + \frac{k \ln n}{n}$$

$$\frac{\mu_4}{(\sigma^2)^2} = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} = E[\mathcal{Z}^4] \quad N(\mu_1 + \beta' (x_2 - \mu_2), \Sigma_{22} - \beta' \Sigma_{21} \beta)$$

$$\hat{\ell}^0(x; \rho) = \rho^*(1 - \rho)^{1-x}, \rho \geq 0 \quad \text{argmin}_{\beta} (y - X\beta)(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j|$$

$$\frac{\Delta x^i}{\Delta y^i} = \pi_{x2} + \frac{\alpha_x e_t}{\alpha_y e_t} + \pi_z \left[\frac{\Delta x_{t-z}}{\Delta y_{t-z}} \right] + \dots + \pi_{\tau} \left[\frac{\Delta x_{t-\tau}}{\Delta y_{t-\tau}} \right] + \left[\eta_{t+\tau} \right]$$

$$\hat{\ell}^0(\rho|x) \propto \rho^*(1 - \rho)^{1-x} \times \frac{\rho^{n-s}(1 - \rho)^{\beta - s}}{B(\alpha, \beta)} \quad E \left[\left(\beta (1 + r_{t+s}) \left(\frac{U(c_{t+s})}{U(c_t)} \right) - 1 \right) z_t \right] = \rho$$

$$\omega = \eta(R\beta - r)' \left[R \hat{\Sigma}_{XX} \hat{S} \hat{\Sigma}_{XX}^{-1} R' \right]^{-1} (R\beta - r) \xrightarrow{d} \chi_m^2 \quad Y_t = \delta + Y_{t-z} + \epsilon_t$$

$$g(e) = \frac{1}{\tau_h} \sum_{t=1}^{\tau} K \left(\frac{e_t - e}{h} \right) \quad l(\lambda, y) = -\delta \lambda + \ln(\lambda) \sum_i y_i - \sum_i \ln(y_i!)$$

$$\hat{\Sigma}^{AW} = \hat{\Gamma}_\theta + \sum_{i=1}^l \frac{l+1-\zeta}{l+1} \left(\hat{\Gamma}_i + \hat{\Gamma}'_i \right) \quad \hat{Y}_t = \beta_{21} X_t + \beta_{22} X_t I_{[X_t > \kappa]} + \epsilon_t \quad \sqrt{\tau} (R(\hat{\theta}) - R(\theta_\rho)) \xrightarrow{d} N \left(0, \frac{\partial R(\theta_\rho)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\rho)}{\partial \theta} \right)$$

$$\rho_s = \frac{\gamma_s}{\gamma_\theta} = \frac{E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])]}{V[y_t]} \quad -2X'y = -2X'\hat{e} = \sigma \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y,x}$$

$$\beta = \frac{\partial l(\theta, y)}{\partial \mu} = \sum_{i=1}^n \frac{(y_i - \mu)}{\sigma^2} \quad A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$$\frac{\partial l(\theta, y)}{\partial \sigma^2} = -\frac{\sigma}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^4} \quad \lambda_{trace}(r) = -\tau \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i)$$

$$\hat{\ell}^0(x_1 | x_2 \in B) = \frac{\int_B \hat{\ell}^0(x_1, x_2) dx_2}{\int_B \hat{\ell}^0(x_2) dx_2} \quad f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$z_t = \Upsilon z_{t-1} + \xi_t \quad \sigma_t^2 = \omega + \alpha V_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$\Sigma_t = CC' + AA' \odot \epsilon_{t-1} \epsilon_{t-1}' + BB' \odot \Sigma_{t-1}$$

$$\text{VaR}_{t+z} = -\mu - \sigma_{t+z} G_{C'}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial l(y; \psi)}{\partial \psi} \frac{\partial l(y; \psi)}{\partial \psi'} \right]$$

$$\sum_{i=1}^n g_i(Y_i) \stackrel{d}{=} E[g(Y)]$$

Expected Shortfall

- Conditional Expected Shortfall (ES, also called Tail VaR)

$$ES_{t+1} = E_t[r_{t+1} | r_{t+1} < -VaR_{t+1}]$$

- "Expected Loss given you have a Value-at-Risk violation"
- Usually requires the specification of a complete model for the conditional distribution
- Uses all of the information in the tail
- Evaluation
 - ▶ Standard Problem, a conditional mean
 - ▶ GMZ regression
 - $H_0 : \gamma = \mathbf{0}$
- Difficult to test since relatively few observations

Coherent Risk Measures

VaR, ES and Density Forecasting



Coherent Risk Measures

- Coherence is a desirable property for a risk measure
 - ▶ But not completely necessary
- ρ is the required capital necessary according to some measure of risk (VaR, ES, Standard Deviation, etc.)
- P, P_1 and P_2 are portfolios of assets
- A Coherent measure satisfies:

Drift Invariance

$$\rho(P + c) = \rho(P) - c$$

Homogeneity

$$\rho(\lambda P) = \lambda \rho(P) \quad \text{for any } \lambda > 0$$

Monotonicity

If P_1 first order stochastically dominates P_2 , then $\rho(P_1) \leq \rho(P_2)$

Subadditivity

$$\rho(P_1 + P_2) \leq \rho(P_1) + \rho(P_2)$$

Coherent Risk Measures

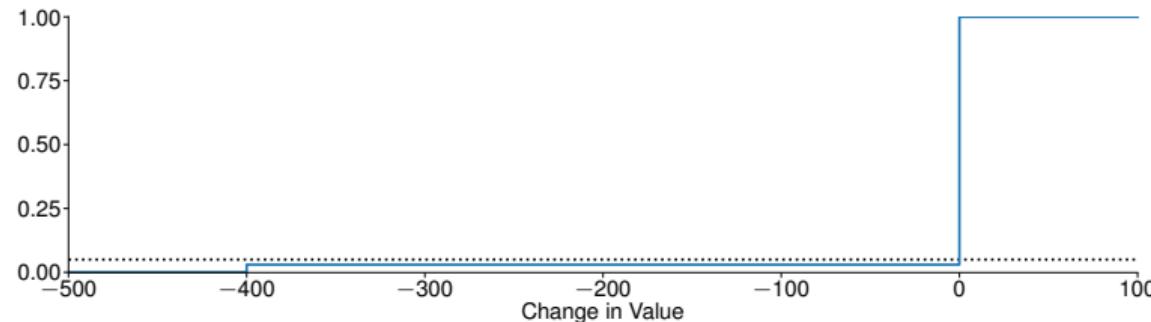
- VaR is *not* coherent
 - ▶ Because VaR is a quantile it may not be subadditive

VaR is Not Coherent

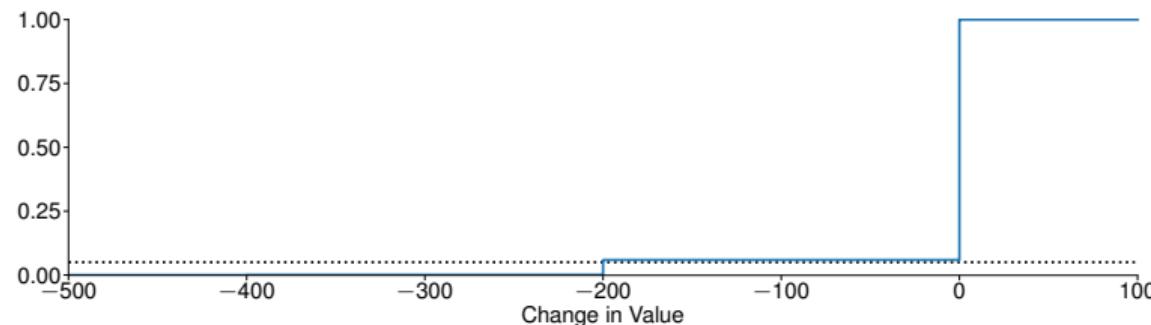
- Two portfolios P_1 and P_2 holding a bond
 - ▶ Each paying 0%, par value of \$1,000
 - ▶ Default probability 3%, recovery rate 60%
 - ▶ Two companies, defaults are independent
- Value-at-Risk of P_1 and P_2 is \$0
- Value-at-Risk of $P_3 = 50\% \times P_1 + 50\% \times P_2 = \200
 - ▶ 5.91% that one or both default

Coherent Risk Measures

P_1 and P_2



P_3



Coherent Risk Measures

- ES is coherent
 - ▶ Doesn't mean much
 - ▶ VaR still has a lot of advantages
 - ▶ More importantly **VaR and ES agree in most realistic settings**

ES is coherent

- ES of P_1 and P_2 is \$240
 - ▶ Given in lower 5% of distribution, 60% chance of a loss of \$400
- ES of P_3
 - ▶ Given in lower 5% of distribution:
 - $0.0009/0.05 = .018$ probability of \$400 loss (2 defaults)
 - $0.0491/0.05 = .982$ probability of \$200 loss (1 default)
 - ES of $\$7.20 + \$196.40 = \$203.60$
- ES is subadditive when VaR is not

Review

Key Concepts

Expected Shortfall, Coherent Risk Measure

Questions

- How does Expected Shortfall differ from VaR?
- How are ES models evaluated for accuracy?
- What is the key limit to the evaluation of ES models?
- What are the four properties of a coherent risk measure?
- Why is VaR not coherent?